AU/507257
PCT/AU03/00300

10 SEP 2004

Patent Office
Canberra

I, SMILJA DRAGOSAVLJEVIC, TEAM LEADER EXAMINATION SUPPORT AND SALES hereby certify that annexed is a true copy of the Provisional specification in connection with Application No. PS 1118 for a patent by PROTEOME SYSTEMS INTELLECTUAL PROPERTY PTY LTD as filed on 13 March 2002.

WITNESS my hand this
Twenty-sixth day of March 2003

S. Dragosavljevic

SMILJA DRAGOSAVLJEVIC
TEAM LEADER EXAMINATION
SUPPORT AND SALES

10/507257

IN THE UNITED STATES RECEIVING OFFICE (RO/US)
Designated/Elected Office (DO/EO/US)

U.S. National Stage of

International Application No.: PCT/AU03/00300

International Filing Date: 13 March 2003

Priority Date Claimed: 13 March 2002

Applicants: Jonathan Wesley Arthur, Marc Wilkins and Mathew Danger Traini

Title: ANNOTATION OF GENOME SEQUENCES

Attorney's Docket No.: 3170.1006-000

Date: 10 September 2004

EXPRESS MAIL LABEL NO. EV214960666US

AUSTRALIA

Patents Act 1990

Proteome Systems Intellectual Property Pty Ltd

PROVISIONAL SPECIFICATION

Invention Title:

Annotation of genome sequences

The invention is described in the following statement:

Field of the Invention

This invention relates to a method of annotation of genome sequences.

Background of the Invention

Many genomes, including the human genome have now been sequenced. A genome sequence provides a list of bases (A,T,G,C) in the order in which they appear in a length of DNA, however, the sequence *per se* tells one very little about the genome that is useful and easily or immediately comprehensible. For example in the study of a disease causing bacteria it would be useful in searching for a cure for the disease to determine the location of that part of the bacterium's genome which expressed a particular protein. However, it can be difficult to predict where proteins of interest may be located in a genome sequence. It cannot always be done simply by looking at the sequence *per se*.

There are a number of known processes for attempting to determine the location of proteins in genome sequence data. One known method uses computer programs to locate recognisable regions such as start codons and stop codons in a DNA sequence. Other programs attempt to locate proteins by locating regions of high complexity within a DNA sequence which typically indicates the location of a protein.

However, these approaches are far from perfect as in order to implement these programs, various assumptions and hypotheses have to be made about the location of a protein of interest in the DNA sequence, in particular, the potential start and stop positions of the protein. A detection method that requires such assumptions or hypotheses may produce incorrect results if the assumptions/hypotheses are incorrect. For example these procedures are unlikely to locate non-typical sequences, which ironically may be of more interest than other proteins having more typical sequences identified using existing techniques.

Thus, it is one object of the present invention to provide a method for annotating genome sequences, which is hypothesis independent and does not make assumptions for the detection of a protein from nucleic acid sequences.

Any discussion of documents, acts, materials, devices, articles or the like which has been included in the present specification is solely for the purpose of providing a context for the present invention. It is not to be taken as an admission that any or all of these matters form part of the prior art base or were

common general knowledge in the field relevant to the present invention as it existed in Australia before the priority date of each claim of this application.

Summary of the Invention

A first broad aspect of the present invention, provides a method of identifying one or more proteins in an unannotated DNA sequence, the method comprising:

(a) dividing the DNA sequence into a plurality of sequence fragments each fragment being of substantially the same length and from about 300 to 5000 base pairs long;

(b) performing a six frame translation of each of the DNA sequence fragments to obtain six translated amino acid sequence fragments for each DNA sequence fragment;

(c) subjecting each of the translated sequence fragments to theoretical digestion to obtain a plurality of cleaved peptide sequences;

(d) comparing experimental empirical data for peptide fragments from a protein digested in the same manner as the theoretical digestion at step (c) with the theoretical data generated in step (c) for each of the translated sequence fragments to identify one or more translated sequence fragments which include a substantial number of peptides present in the digested protein.

An advantage of the present invention is that no assumptions need to be made about the location of proteins in the DNA sequence data. DNA sequences with non-typical stop and or start codons may be located. The results are hypothesis independent.

Typically the theoretically generated peptide masses are compared to the masses of the peptides experimentally generated by the digested protein and the sequence fragment which has the greatest number of theoretical peptide masses correlating to the empirical data indicates the likely location of the protein of interest in the DNA sequence. The masses of the peptides experimentally generated from the digested protein will typically be determined by mass spectrometry.

It is preferred that the DNA sequence is duplicated and the original and duplicate are split in such a manner that the sequence fragments from the original overlap the cuts in the original genome sequence.

It is important that the sequence fragments are approximately the same length as one another and are sized to equate to the length of a typical protein.

Hence, each fragment is, as discussed above, about 300-5000 base pairs long. Proteins vary in size, most proteins being 10 to 100 kDa i.e. about 300-3000 base pairs long. Most preferably, the sequence fragments will be around 1000 or 1050 bases long, the latter translating to 350 amino acids which is approximately equivalent to a 33 to 37 kDa protein, which is a common size for a protein.

Using DNA sequences of approximately that length produce about 12 to 20 hits against a background number of hits of around 4 for sequences which do not contain a protein.

In a related aspect of the present invention, the step of dividing the DNA sequence and the step of performing the six frame translation can be reversed. Hence, a second broad aspect of the present invention provides a method of identifying one or more proteins in unannotated DNA sequence, the method comprising:

(a) performing a six frame translation of a DNA sequence to provide six translated amino acid sequences;

(b) dividing the six translated amino acid sequences into a plurality of fragments, each fragment comprising 100-1666 amino acids;

(c) subjecting each of the fragments to theoretical digestion to obtain a plurality of cleaved peptide sequences;

(d) comparing experimental empirical data for peptide fragment for peptide fragments from a protein digested in the same manner as the theoretical digestion at step (c) with theoretical data generated in step (c) for each of the fragments to identify one or more fragments which include a substantial number of peptides present in the empirically digested protein.

Brief Description of the Drawings

A specific embodiment of the present invention will now be described by way of example with reference to the accompanying drawings in which:

Figures 1A to 1E are schematic diagrams illustrating various steps in the method of the present invention;

Figure 2 shows a report comparing empirical masses of the protein Glycerol-3-phosphate dehydrogenase (P95113) against a split, translated, and digested genome sequence of M Tuberculosis;

Figure 3 shows a detailed report for the frame which provided the greatest number of hits (eighteen) comparing the theoretical masses of each peptide in the digest with the empirical masses; and

Figure 4 shows detailed reports for two frames which have provided only four hits;

Figure 5 shows detailed results for one frame/fragment number 6366; and

Figure 6 shows detailed results for another frame/fragment number 6364.

Detailed Description of a Preferred Embodiment

Referring to the drawings, Figure 1A, shows a genome sequence 10 which is taken and split into a series of shorter genome sequences or sequence fragments 12. Overlapping sequences are preferably provided by duplicating the genome sequence and cleaving the duplicated sequence at locations midway between the breaks in the original sequence so that the sequences (12a,12b..., 14a, 14b...) are overlapping as shown in Figure 1A.

Typically, the genome will be cut into sequence fragments which are 1050 bases long. This approximates to 350 amino acids which will be found in a protein of around 33 to 37 kDa which is a common protein size. A bacterium such as Mycobacterium tuberculosis (Tb) will have around 4.4 million bases in its genome. Duplicating and cutting that genome will result in approximately 8400 sequence fragments.

A six frame translation is then carried out on each of the sequence fragments. Figure 1B schematically illustrates a 6 frame translation carried out on one of the sequence fragments (14d). For each fragment, six virtual proteins are produced. Fragment 14d produces six virtual proteins 16a-16g. Using the M Tuberculosis example referred to above the 8400 virtual proteins become 50,400. These virtual proteins are then subjected to theoretical digestion according to rules which mimic the action of an endoproteinase enzyme such as trypsin which cut at specific target sites on a target sequence. This digestion is schematically illustrated in Figure 1C. Each virtual protein becomes a series of "virtual peptides" and the mass of each virtual peptide is calculated. "Protein" 16g becomes six peptides 18a to 18g. Fewer or more peptides may be produced from each virtual protein. The protein of interest is then subjected to an empirical digestion using the same enzyme and peptide

mass data is obtained from mass spectrometry of the peptides expressed by that protein.

The masses of the various empirically derived peptides are then compared with the theoretical peptide masses produced by theoretical cleavage of the sequence fragments. This is done in a stepwise manner and frame by frame whereby all the empirical peptide masses are matched against all peptides from the first virtual protein and the number of matching peptides (hits) is recorded. For each virtual protein, this process is carried out six times, once for each of the amino acid translations. However, the number of hits for each frame is calculated separately and the hits are not summed together. This process is then repeated for the second virtual protein and so on, until it has been carried out for all the virtual proteins. This step is illustrated in Figure 1D. There is a background number of hits. Typically, each theoretical protein or sequence fragment will produce about 3 or 4 peptides having masses which correlate to masses produced by the actual empirical digest of the protein of interest. The sequence fragment which produced the protein of interest will in contrast typically have 12 to 20 peptide matches with the empirical digest of the protein of interest but is limited by the number of peptides generated empirically.

Clearly the relevant part of the genome sequence may have been cut in the original division of the genome sequence, however the overlapping of the original and duplicate genome sequences reduces the risk of this. In any case even if the protein is split it is still possible to identify the relevant part of the genome sequence because we would see a reasonable number e.g. 6 to 10 hits in two adjacent overlapping fragments. The part of the sequence which carries the most peptide masses which match the peptide masses produced by the empirical digestion and has a number of hits which is clearly above the background (noise) level is likely to be that part of the genome which carries the protein of interest. By knowing where the part of the sequence came from, this identifies the location of the protein in the genome sequence (Figure 1E).

The present invention works particularly well with small genomes such as bacterial and yeast genomes or other eukaryote genomes that have few introns and small amounts of non-coding DNA.

The method can also be used for the detection of pseudo genes which are versions of genes which have become defunct and identifying "protein families" of similar proteins. When a protein from a family of proteins is

detected, a number of regions having a large number of matches may be identified. This indicates that the proteins may be members of the same protein family which may be for example be expressed in different tissues. The method also identifies if there is a now inactive or defunct family region of genome to a related region that carries an active version of a protein.

Example

Figures 2 to 6 illustrate the results of carrying out the method of the present invention looking for the part of the M Tuberculosis genome which produces the protein Glycerol-3-phosphate dehydrogenase (P95113). The protein was digested with trypsin and the masses of the peptides produced by the digestion measured using mass spectometric analysis.

The empirical masses were searched against the split, translated and theoretically digested genome sequence for M Tuberculosis using the method of the present invention as discussed above. Figure 2 shows a summary of the results illustrating all the theoretical sequence fragments which produced four hits or more. One fragment 6365 derived from the genome base pairs 3341051 to 3342100 produced 17 hits. This fragment contains the complete sequence for the protein of interest. Two adjacent fragments 6364 and 6366 produced eight hits each, as they are overlapping and contain part of the protein of interest. Other results show a background number of 4 or 5 hits only.

Figure 3 is a detailed report on segment 6359. For each peptide in the digest, the theoretical database mass is compared with the empirical mass determined by mass spectrometry. The start and end positions of each of the peptides in the sequence are also given. Note that in Figure 2, the peptides are ordered by increasing mass. They could simply be ordered by position in the sequence. The report also indicates if any cleavages have been missed (MC) and the frame number "5" indicates that the fifth frame (out of the six frame translations of the genome fragment of interest) produced the match with the empirical data.

The sequence is shown at 50 at the top of the report. Note that there are no stop codons 52 marked with an asterisk in the areas where the protein is located.

In contrast, Figure 4 shows the reports on two of the fragments (numbers 238 and 749) which had four hits only - a background number of hits. The matching peptides are scattered throughout the sequence. In the case of

fragment 238 there are frequent stop codons suggesting that this is not a coding region of the genome. For fragment 749 there is a background number of hits to what is probably a coding region of the genome, but coding a protein different to the one we are analysing.

Figure 5 shows detailed results for fragment number 6366 which had 8 hits. This shows that a portion of the protein sequence is present in the second half of this fragment. Figure 6 shows detailed results for fragment number 6364. this shows that a portion of the protein sequence is present in the first half of this fragment.

It will be appreciated by persons skilled in the art that numerous variations and/or modifications may be made to the invention as shown in the specific embodiments without departing from the spirit or scope of the invention as broadly described. The present embodiments are, therefore, to be considered in all respects as illustrative and not restrictive.

Dated this thirteenth day of March 2002

Proteome Systems Intellectual
Property Pty Ltd
Patent Attorneys for the Applicant:

F B RICE & CO

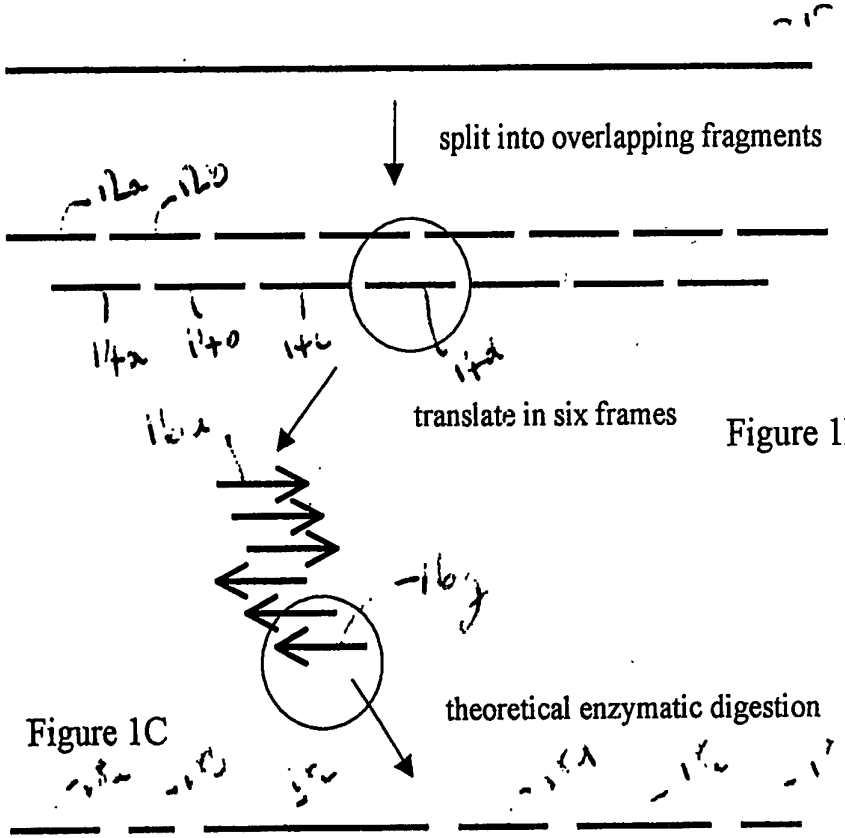


Figure 1A

Figure 1B

Figure 1C

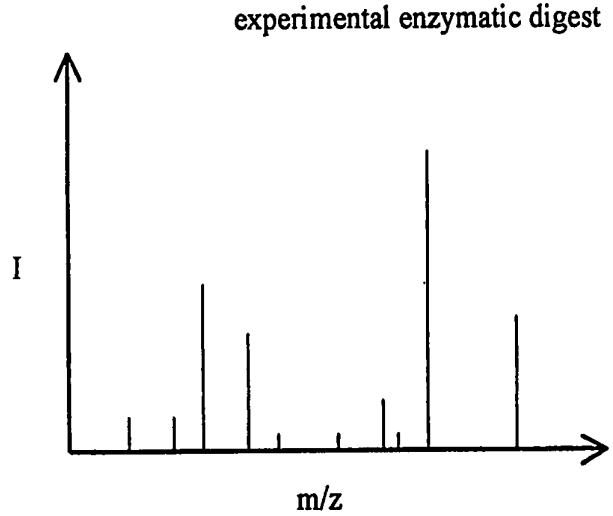


Figure 1D

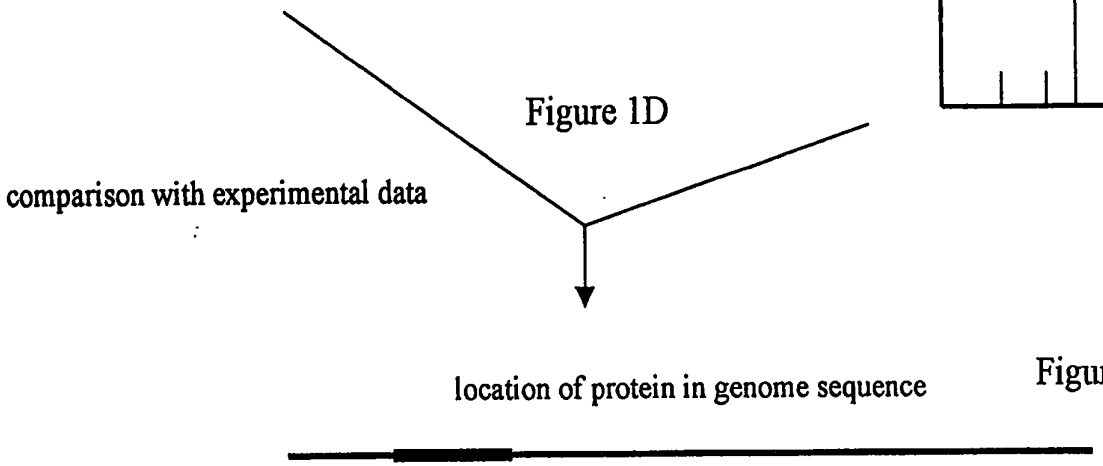


Figure 1E

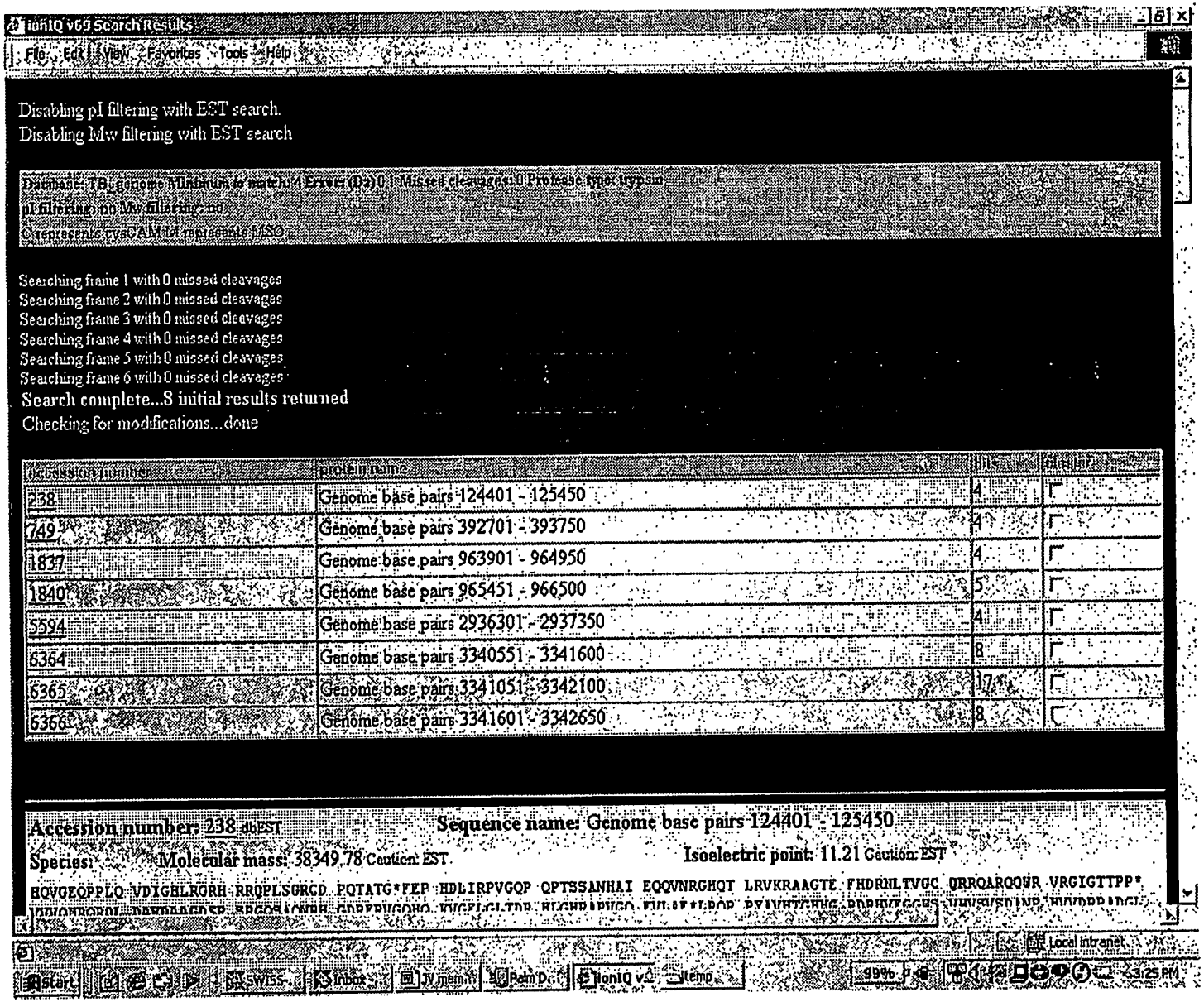

IonIQ v69 Search Results
File Edit View Favorites Tools Help
Disabling pI filtering with EST search.
Disabling Mw filtering with EST search
Searching frame 1 with 0 missed cleavages
Searching frame 2 with 0 missed cleavages
Searching frame 3 with 0 missed cleavages
Searching frame 4 with 0 missed cleavages
Searching frame 5 with 0 missed cleavages
Searching frame 6 with 0 missed cleavages
Search complete...8 initial results returned
Checking for modifications...done
accession number
protein name
hits
238
Genome base pairs 124401 - 125450
4
749
Genome base pairs 392701 - 393750
4
1837
Genome base pairs 963901 - 964950
4
1840
Genome base pairs 965451 - 966500
5
5594
Genome base pairs 2936301 - 2937350
4
6364
Genome base pairs 3340551 - 3341600
8
6365
Genome base pairs 3341051 - 3342100
17
6366
Genome base pairs 3341601 - 3342650
8
Accession number: 238 dbEST
Sequence name: Genome base pairs 124401 - 125450
Species:
Molecular mass: 38349.78 Caution: EST
Isoelectric point: 11.21 Caution: EST
HQVGEQPPLQ VDIGHLRGRH RRQPLSGRCD PQTATG*FEP HDLIRPVGQP QPTSSANHAI EQQVNRGHQT LRVKRAAGTE FHDRNLTVGC QRRQARQQUR VRGIGTTPP*
Local intranet
Start
99%
3:25 PM

Figure 2

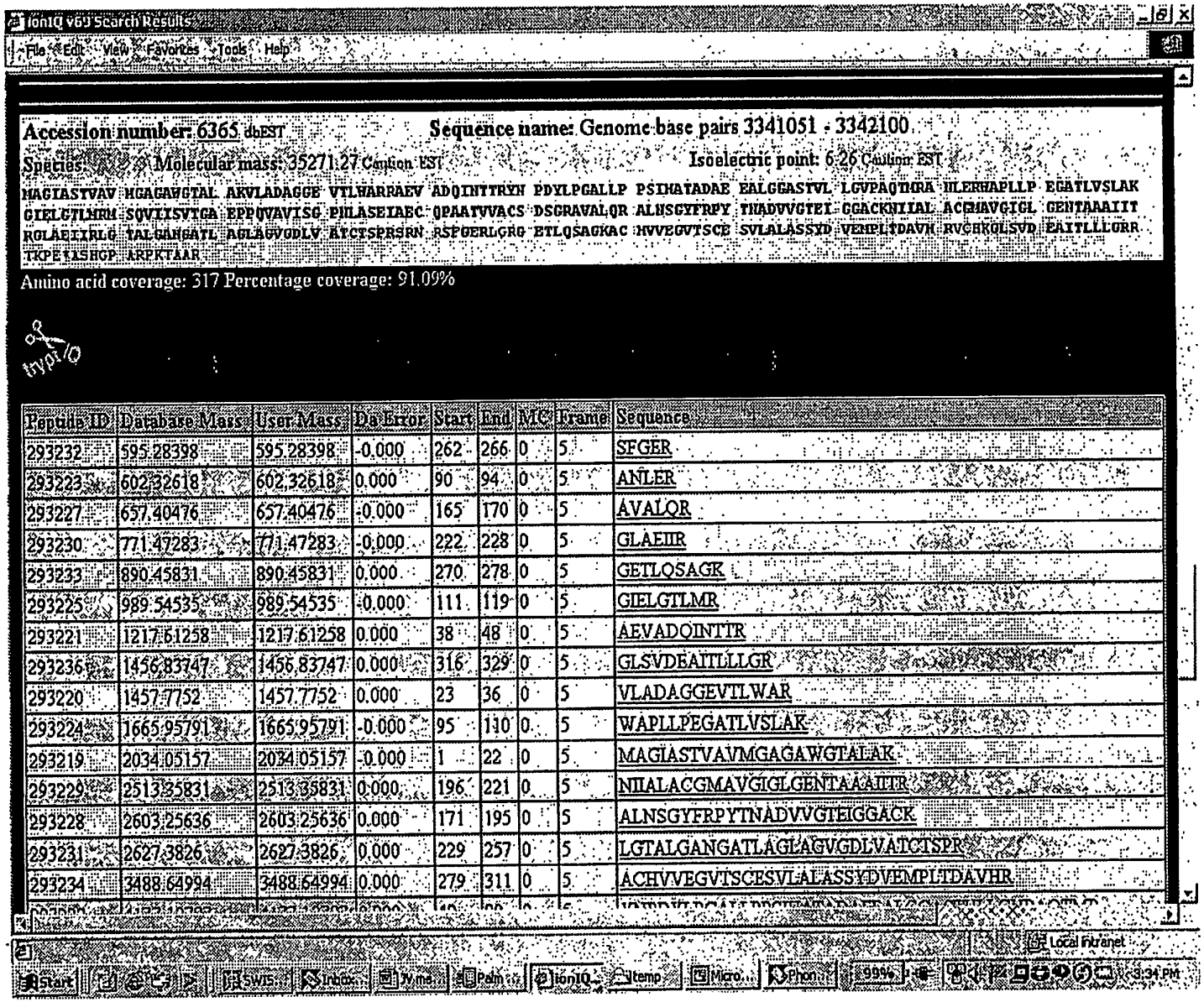

IonIQ v6J Search Results
File Edit View Favorites Tools Help
Accession number: 6365 dbEST
Sequence name: Genome base pairs 3341051 - 3342100
Species: Molecular mass: 35271.27 Caution: EST
Isoelectric point: 6.26 Caution: EST
MAGIASTVAV MGAGAWGTAL AKVLADAGGE VTLWARNAEV ADQINTTRYN PDYLPGALLP PSIHATADAE EALGGASTVL LGVPAQTMRA ILERWAPLLP EGATLVSLAK
GIELGTLMRM SQVIISVTGA EPPQVAVISG PILASEIAEC QPAATVVACS DSGRAVALQR ALNSGYFRPY TNADVVGTEI CGACKNIIAL ACGMAVGIGL GENTAAAIIT
RGLAEIIRLG TALGANGATL AGLAGVGDLV ATCISPRSRN RSFGERLGRG ETLQSAGKAC HVVEGVTSCE SVLALASSYD VEMPLITDAV HRVCHKGLSV DEAITLLLGR
TKPEIISHGP ARPKTAAR
Amino acid coverage: 317 Percentage coverage: 91.09%
Peptide ID
Database Mass
User Mass
Da Error
Start
End
MC
Frame
Sequence
293232 595.28398 595.28398 -0.000 262 266 0 5 SFGER
293223 602.32618 602.32618 0.000 90 94 0 5 ANLER
293227 657.40476 657.40476 -0.000 165 170 0 5 AVALQR
293230 771.47283 771.47283 -0.000 222 228 0 5 GLAEIIR
293233 890.45831 890.45831 0.000 270 278 0 5 GETLQSAGK
293225 989.54535 989.54535 -0.000 111 119 0 5 GIELGTLMR
293221 1217.61258 1217.61258 0.000 38 48 0 5 AEVADQINTTR
293236 1456.83747 1456.83747 0.000 316 329 0 5 GLSVDEAITLLLGR
293220 1457.7752 1457.7752 0.000 23 36 0 5 VLADAGGEVTLWAR
293224 1665.95791 1665.95791 -0.000 95 110 0 5 WAPLLPEGATLVSLAK
293219 2034.05157 2034.05157 -0.000 1 22 0 5 MAGIASTVAVMGAGAWGTALAK
293229 2513.35831 2513.35831 0.000 196 221 0 5 NIIALACGMAVGIGLGENTAAAIITR
293228 2603.25636 2603.25636 0.000 171 195 0 5 ALNSGYFRPYTNADVVGTEIGGACK
293231 2627.3826 2627.3826 0.000 229 257 0 5 LGTALGANGATLAGLAGVGDLVATCISPR
293234 3488.64994 3488.64994 0.000 279 311 0 5 ACHVVEGVTSCESVLALASSYDVEMPLITDAVHR
Local intranet
Start
3:34 PM

Figure 3

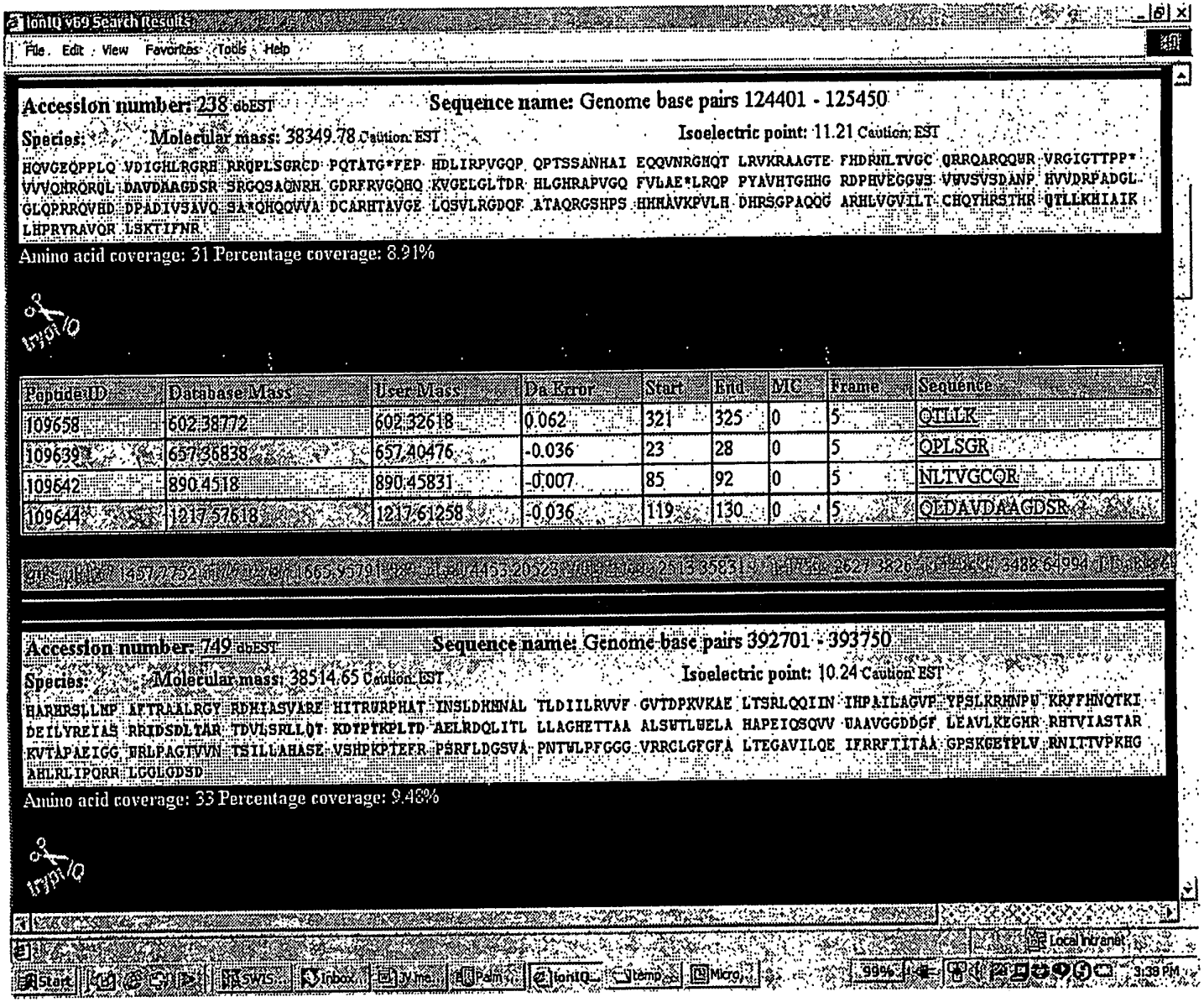

IonIQ v69 Search Results
File Edit View Favorites Tools Help
Accession number: 238 dbEST
Sequence name: Genome base pairs 124401 - 125450
Species:
Molecular mass: 38349.78 Caution: EST
Isoelectric point: 11.21 Caution: EST
HQVGEQPPLQ VDIGHLRGRH RRQPLSGRCD PQTATG*FEP HDLIRPVGQP QPTSSANHAI EQQVNRGHQT LRVKRAAGTE FHDRNLTVGC QRRQARQQUR VRGIGTTPP*
VVVQHRORQL DAVDAAGDSR SRGQSACNRH GDRFRVGQHQ KVGELGLTDR HLGHRAPVGQ FVLAE*LRQP PYAVHTGHHG RDPHVEGGVS VHVSVSDANP HVVDRFADGL
GLQPRRQVHD DPADIVSAVQ SA*QHQQVVA DCARHTAVGE LQSVLRGDQF ATAQRGSHPS HHHAVKPVLH DHRSGPAQQG ARHLVGVILT CHQYHRSTHR QTLLKHIAIK
LHPRYRAVQR LSKTIFNR
Amino acid coverage: 31 Percentage coverage: 8.91%
Trypsin
Peptide ID
Database Mass
User Mass
Da Error
Start
End
MC
Frame
Sequence
109638
602.38772
602.32618
0.062
321
325
0
5
QTLLK
109639
657.36838
657.40476
-0.036
23
28
0
5
QPLSGR
109642
890.4518
890.45831
-0.007
85
92
0
5
NLTVGCQR
109644
1217.57618
1217.61258
-0.036
119
130
0
5
QLDAVDAAGDSR
Accession number: 749 dbEST
Sequence name: Genome base pairs 392701 - 393750
Species:
Molecular mass: 38514.65 Caution: EST
Isoelectric point: 10.24 Caution: EST
HARHRSLLMP AFTRAALRGY RDHIASVARE HITRVRPHAT INSLDKMNAL TLDIILRVVF GVTDPKVKAE LTSRLQQIIN IHPAILAGVF YPSLKRHNPV KRYFHNQTKI
DEILYREIAS RRIDSDLTAR TDVLSRLLQT KDTPTKPLTD AELRDQLITL LLAGHETTAA ALSWTLWELA HAPEIQSQVV UAAVGGDDGF LEAVLKEGHR RHTVIASTAR
KVTAPAEIGG HRLPAGTVVN TSILLAHASE VSHPKPTEFR PSRFLDGSVA PNTWLPFGGG VRRCLGFGFA LTEGAVILQE IFRRFTITAA GPSKGETPLV RNITTVPKHG
AHLRLIPORR LGGLGDSD
Amino acid coverage: 33 Percentage coverage: 9.48%
Trypsin
Local intranet
Start
SWIS...
Inbox...
Pdm...
IonIQ
Temp...
Microg...
99%
3:39 PM

Figure 4

4/6

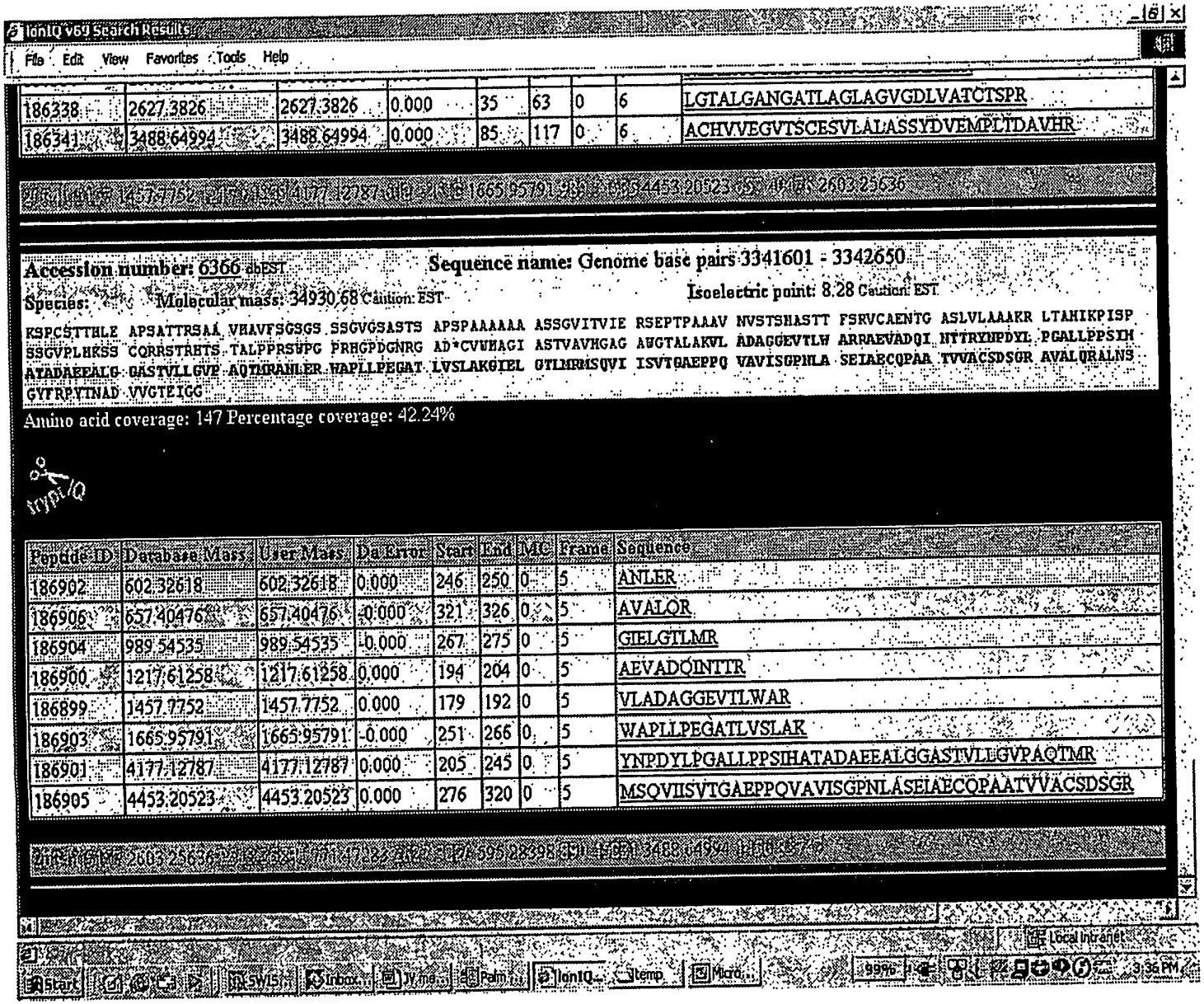
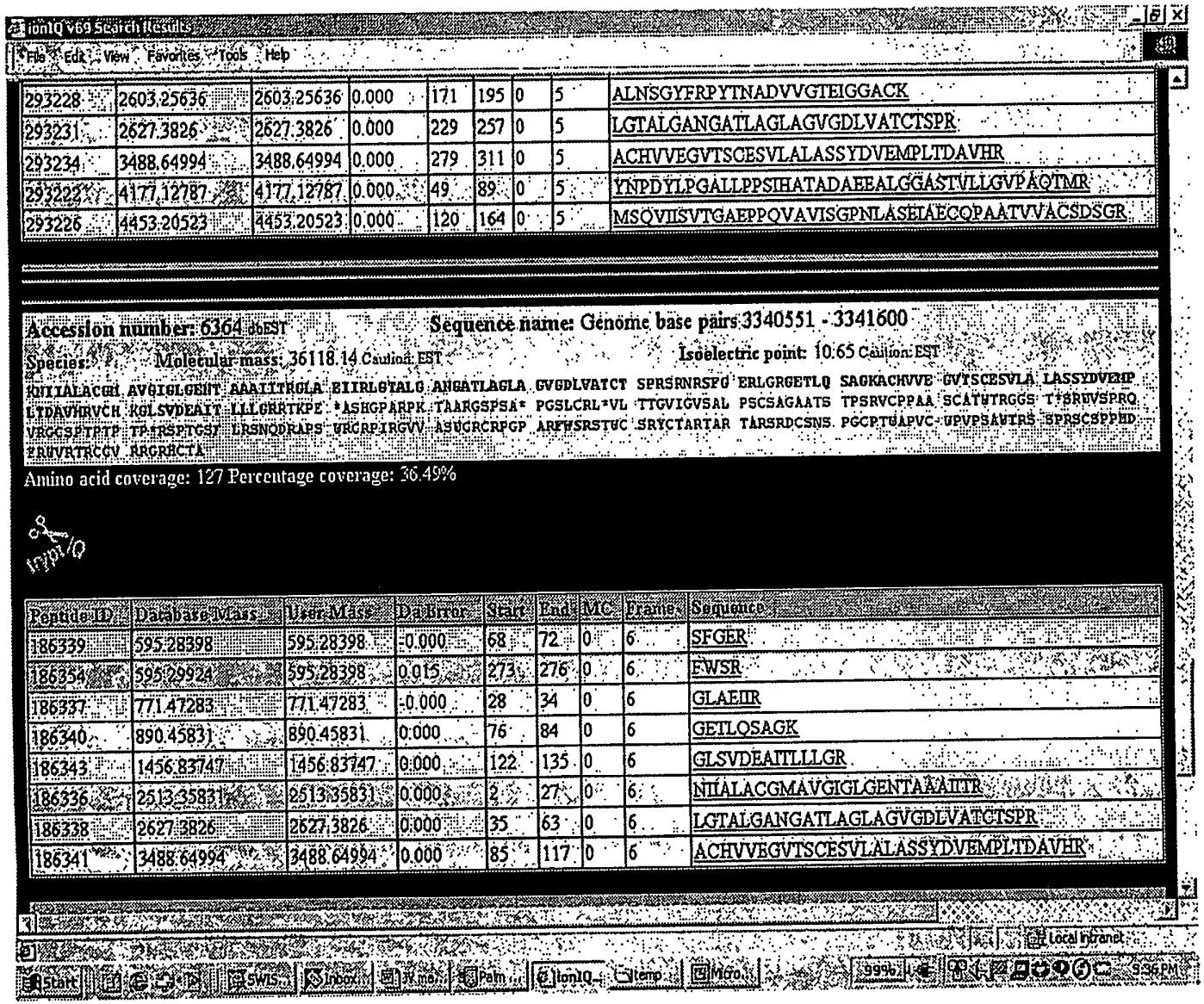

IonIQ v69 Search Results
File Edit View Favorites Tools Help
186338 2627.3826 2627.3826 0.000 35 63 0 6 LGTALGANGATLAGLAGVGDLVATCTSPR
186341 3488.64994 3488.64994 0.000 85 117 0 6 ACHVVEGVTSCESVLALASSYDVEMPLTDAVHR
Accession number: 6366 dbEST
Sequence name: Genome base pairs 3341601 - 3342650
Species:
Molecular mass: 34930.68 Caution: EST
Isoelectric point: 8.28 Caution: EST
KSPCSTTHLE APSATTRSAA VHAVFSGSGS SSGVGSASTS APSPAAAAAA ASSGVITVIE RSEPTPAAAV NVSTSHASTT FSRVCAENTG ASLVLAAAKR LTAHIKPISP SSGVPLHKSS CQRRSTRHTS TALPPRSVPG PRHGPDGNRG AD*CVVHAGI ASTVAVHGAG AVGTALAKVL ADAGGEVTLW ARRAEVADQI HTTRYNPDYL PGALLPPSIH ATADAEEALG GASTVLLGVP AQTMRANLER WAPLLPEGAT LVSLAKGIEL GTLMRMSQVI ISVTGAEPPQ VAVISGPNLA SEIAECQPAA TVVACSDSGR AVALQRALNS GYFRPVTNAD VVGTEIGG
Amino acid coverage: 147 Percentage coverage: 42.24%
trypIQ
Peptide ID Database Mass User Mass Da Error Start End MC Frame Sequence
186902 602.32618 602.32618 0.000 246 250 0 5 ANLER
186906 657.40476 657.40476 -0.000 321 326 0 5 AVALQR
186904 989.54535 989.54535 -0.000 267 275 0 5 GIELGTLMR
186900 1217.61258 1217.61258 0.000 194 204 0 5 AEVADQINTTR
186899 1457.7752 1457.7752 0.000 179 192 0 5 VLADAGGEVTLWAR
186903 1665.95791 1665.95791 -0.000 251 266 0 5 WAPLLPEGATLVSLAK
186901 4177.12787 4177.12787 0.000 205 245 0 5 YNPDYLPGALLPPSIHATADAEEALGGASTVLLGVPAQTMR
186905 4453.20523 4453.20523 0.000 276 320 0 5 MSQVIISVTGAEPPQVAVISGPNLASEIAECQPAATVVACSDSGR
Local intranet
Start
99%
3:36 PM

Figure 5


IonIQ V69 Search Results
File Edit View Favorites Tools Help
293228 2603.25636 2603.25636 0.000 171 195 0 5 ALNSGYFRPYTNADVVGTEIGGACK
293231 2627.3826 2627.3826 0.000 229 257 0 5 LGTALGANGATLAGLAGVGDLVATCTSPR
293234 3488.64994 3488.64994 0.000 279 311 0 5 ACHVVEGVTSCESVLALASSYDVEMPLTDAVHR
293222 4177.12787 4177.12787 0.000 49 89 0 5 YNPDYLPGALLPPSIHATADAEEALGGASTVLLGVPAQTMR
293226 4453.20523 4453.20523 0.000 120 164 0 5 MSQVIISVTGAEPPQVAVISGPNLASEIAECQPAATVVACSDSGR
Accession number: 6364 dbEST
Sequence name: Genome base pairs 3340551 - 3341600
Species:
Molecular mass: 36118.14 Caution: EST
Isoelectric point: 10.65 Caution: EST
QIIIALACGI AVGIGLGENT AAAITRGLA EIIRLGTALG ANGATLAGLA GVGDLVATCT SPRSRNRSFG ERLGRGETLQ SAGKACHVVE GVTSCESVLA LASSYDVEMP LTDAVHRVCH KGLSVDEAIT LLLGRRTKPE *ASHGPARPK TAARGSPSA* PGSLCRL*VL TTGVIGVSAL PSCSAGAATS TPSRVCPPAA SCATVTRGGS T*SRDVSPRQ VRGCSPTPTP TP*RSPTGSY LRSNQDRAPS VRGRPIRGVV ASVGRCRPGP ARFWSRSTGC SRYCTARTAR TARSRDCSNS PGCPTGAPVC UPVPSAVTRS SPRSCSPPHD ERDVRTRCGV RRGRHCTA
Amino acid coverage: 127 Percentage coverage: 36.49%
Input/O
Peptide ID Database Mass User Mass Da Error Start End MC Frame Sequence
186339 595.28398 595.28398 -0.000 68 72 0 6 SFGER
186354 595.29924 595.28398 0.015 273 276 0 6 EWSR
186337 771.47283 771.47283 -0.000 28 34 0 6 GLAEIIR
186340 890.45831 890.45831 0.000 76 84 0 6 GETLQSAGK
186343 1456.83747 1456.83747 0.000 122 135 0 6 GLSVDEAITLLLGR
186336 2513.35831 2513.35831 0.000 2 27 0 6 NIIALACGMAVGIGLGENTAAAITR
186338 2627.3826 2627.3826 0.000 35 63 0 6 LGTALGANGATLAGLAGVGDLVATCTSPR
186341 3488.64994 3488.64994 0.000 85 117 0 6 ACHVVEGVTSCESVLALASSYDVEMPLTDAVHR
Local intranet
Start
SWIS...
Inbox
Pam...
IonIQ
Temp
Micro...
99%
9:36 PM

6/6

Figure 6